**Exploring Movie and Web series Choices with IMDB**

21l-5657 M. Talha Imran

21l-5625 Rohan Javed

21l-5651 M. Mashhood Husnain

**Project Description:**

Our project aims to explore and analyze user preferences regarding movies and web series by using the extensive IMDb database. With so many streaming platforms and lots of shows and movies to pick from, it’s important to know what makes people interested. This helps the people who make, share, and enjoy this content. By applying data analysis and visualization techniques we will see what people really enjoy in different types of shows and movies and using that to predict how successful any newly released movie or show will be.

**Project Phases**

**Data Scraping:**

We will be using Beautiful Soup to extract rating, genre, reviews, actors starred, release date, directors, etc.

**Data Wrangling:**

We will clean and preprocess the scraped data to address issues like missing values, duplicates, and inconsistencies, ensuring data quality and integrity.

**Feature Extraction:**

Employing techniques like sentiment analysis on reviews, genre analysis, actor/actress popularity, and other feature extraction techniques for further processing.

**Machine Learning Models:**

We'll apply different machine learning models and compare them with each other to seek the most accurate predictions. Models like regression and classification, to identify patterns and correlations within the dataset. BOW, TF-IDF, Support vector machines (SVM), and relational neural networks (RNN) will also be used for the sentimental analysis.

**Evaluation:**

Assessing model performance using relevant metrics like accuracy, precision, recall, and F1-score. Additionally, constructing confusion matrices to provide a more detailed evaluation.

**Visualization of Results:**

Using the Plotly and matplotlib libraries, we will create interactive visualizations, showcasing model performance, feature importance, and key insights obtained from the analysis. This way we can present a detailed and engaging picture of the results.